Cryptocurrency Sentiment Analysis Platform Proposal

Team Name: Team Codes-A-Lot
Github Repository: Crypto-Currency-Sentiment-Analysis
Team Members:

Isaac Way
Calvin Choe
Savannah Norem
Nick Greene

Introduction

Our proposition is to create a platform for analyzing online sentiment regarding cryptocurrencies over time. The cryptocurrency market is currently booming and redefining the way finance is handled across the world. The price of most cryptocurrency coins is incredibly volatile when compared to traditional financial constructs. The volatility means that there is substantial risk to lose money when holding cryptocurrency, but also great opportunity to make money. Hence, if one could find a way to determine when cryptocurrency prices were getting ready to rise, then one could make a profit. Since cryptocurrencies are priced entirely on perceived values, it makes sense that public sentiment would be correlated with price on some level. This project aims to gather sentiment data by scraping the web, and then package that data so that it might be analyzed in a useful way.

Team Codes-A-Lot is planning to experiment with text analysis and word association to analyze changes in the public perception of cryptocurrency, and to determine if the public sentiment can be used to predict short-term rises and falls in the market values. We will be using standard text-analysis tools as well as machine-learning algorithms and scikit-learn's neural network.

There are commercially available products that perform this sort of duty with varying degrees of success, however, the team is not aware of any open-source projects that attempt anything this ambitious. We hope to produce a product that is both flexible and accurate.

Our team includes four members, three of whom are current Computer Science undergraduates, as well as a University of Tennessee graduate student with an accounting degree. All members are experienced programmers with backgrounds in multiple programming languages. By including an accounting graduate, we're given an entirely different perspective on the project. Combining all our skillsets will allow us to produce an effective product.

Customer Value

The potential customers for this project are any person who has money invested in a cryptocurrency that our platform offers analysis for. Since the objective of investing is to make money, anyone invested in cryptocurrency wants just that: to make money. We aim to provide an interface that provides sentiment based information of a cryptocurrency that can be used as a tool

to make more educated decisions when buying and selling cryptocurrency. Furthermore, by showing historical sentiment concurrently with historical crypto pricing data the platform will provide insight as to HOW sentiment typically affects the cryptocurrency markets. This will allow users to learn about the typical correlation between sentiment and prices, which provides means for a customer to become a more informed investor, hopefully allowing them to make more money through crypto investing.

There are multiple ways that the success of our platform might be measured. The most obvious way is if a user uses the tool and improves their crypto-folio more than they would have if the user had not used the platform. This would be a hard thing to measure effectively for multiple reasons. Firstly, it depends fully on the user using the tool to make real and accurate insights. If one customer is a bad investor to begin with, they might report bad results regardless of the actual effectiveness of our platform. Furthermore, crypto markets are notoriously volatile and seem to shoot up and down almost randomly at times, so even if a savvy-investor user uses our tool in a way that makes a lot of sense it's still possible they might take on losses. Hence, for development purposes the best way to measure success might be to use the tool ourselves and gauge whether it is providing us with meaningful cryptocurrency insights. By running our tool on historical data and determining whether the sentiment analysis data we create provides a correlation with the cryptocurrencies actual price. If we can create a system that consistently returns positive sentiment over periods of time leading up to an increase in price, then we know our system is providing meaningful insight.

Many commercial sentiment analysis platforms for cryptocurrency already exist. A simple google search for "Cryptocurrency sentiment analysis" yields many different results. However, all of them are subscription based. So if we were to make our project open source we would be creating something that does not yet exist on the market.

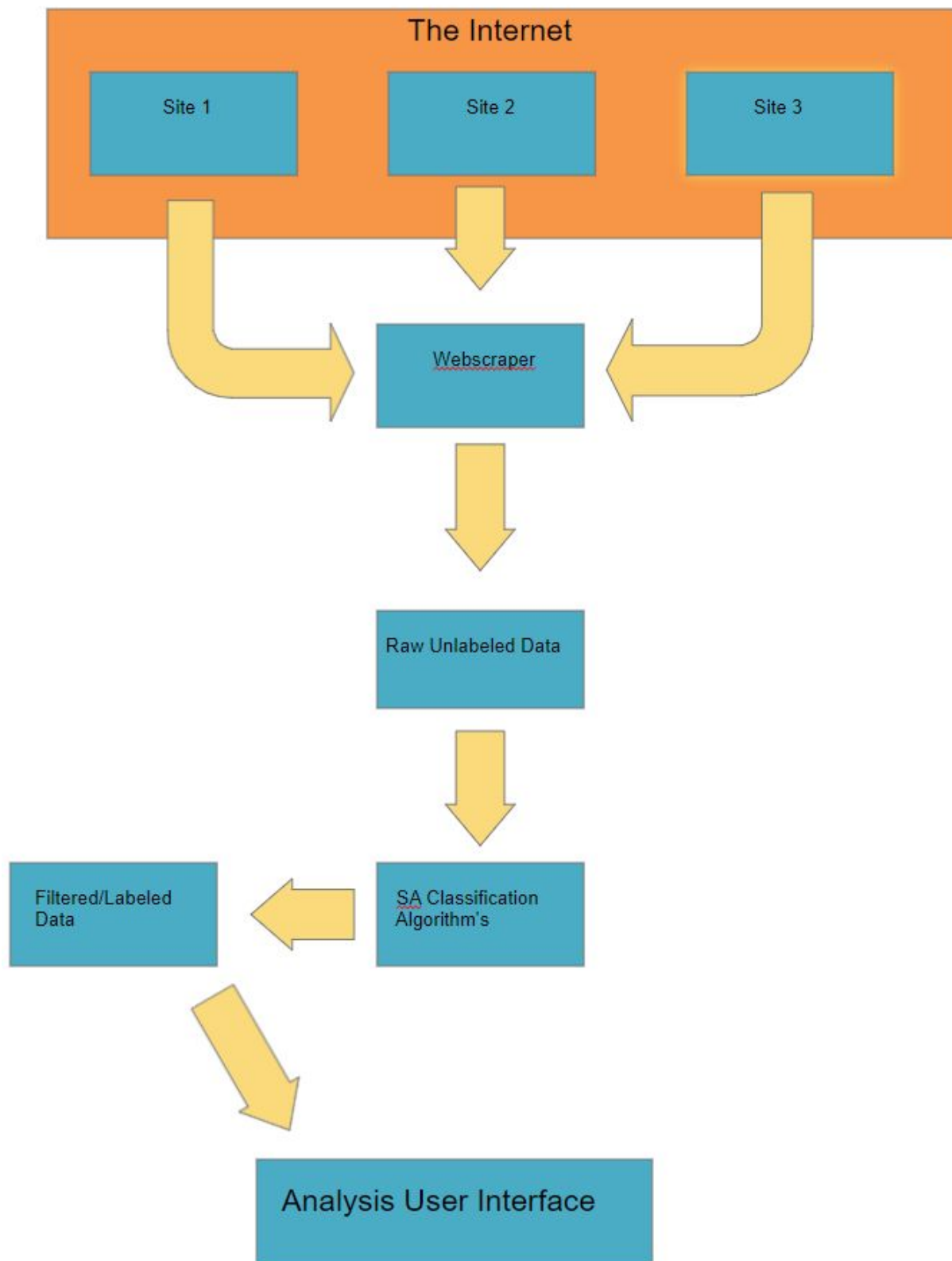<center>**Solution and Technology**</center>

At the least our system needs to accomplish 3 seperate things. First, the software needs to be able to acquire sentiment articles as well as cryptocurrency prices. Acquisition will be accomplished by scraping websites that have this data, such as search engines and social media websites. To create this part of the platform we will use freely available web scraping frameworks. Specifically, Scrapy for sites that don't require browser interaction, and Selenium for sites that do. By leveraging these frameworks we can easily write scripts to acquire any data we might want and save it for use in other parts of the algorithm. To acquire crypto prices, we will get price data directly from cryptocurrency exchanges, each of which typically provide an Python API for getting prices.

The second piece is the sentiment analysis algorithm itself. The algorithm needs to take as input a string of text and label it as having positive or negative sentiment. Since we only care about positive and negative sentiment, we will need to implement a second algorithm that

determines whether or not an article is neutral so that we might filter out neutral articles automatically. To create these algorithms we turn to machine learning classification algorithms. There are a vast number of ML libraries available freely that we might use. Within the current code base we have implemented both of these algorithms using scikit-learn's neural network. In its current state the positive/negative algorithm is performing very well while the neutrality algorithm performs somewhat-okay but hopefully it can be improved further. Obviously we want to achieve as high performing a model as is possible since accurate analysis of sentiment data requires that the sentiment data itself is accurate, so other options of algorithms will be explored as well as methods to improve those algorithms.

Last is the actual user interface for viewing sentiment analysis and pricing data. The interface should allow the user to view historical prices of cryptocurrencies within specified time frames and view the level of sentiment for that same time period. The exact frameworks for creating this interface have not yet been decided, however node.js looks like one promising option as it would allow the developers a straightforward route to creating a UI that allows for showing data through graphs and such.

After completing the basic requirements, we will be able to deliver some value to the customers because we're allowing them to view sentiment analysis data alongside crypto prices, which is an useful investment tool as long as the sentiment data is accurate. The platform might be further improved by adding more tools for analyzing sentiment data. One such way to do this would be by adding a feature to calculate and show Technical Analysis indicators, such as the moving average or the MACD, alongside the pricing and sentiment data. Technical analysis could be incorporated with both sentiment numerics as well as pricing numerics in order to create meaningful displays, and could be accomplished with any of the multiple Technical Analysis libraries that are available for Javascript or even Python. To test our system, the team will test on a feature by feature basis. By making sure each piece of the project is functioning on an incremental basis, we can make sure the entire project functions as a whole.

The Internet

| Site 1 | Site 2 | Site 3 |

Webscraper

Raw Unlabeled Data

Filtered/Labeled Data ← SA Classification Algorithm's

Analysis User Interface

Team

Previously, two members of the current team (Calvin and Isaac) have created the beginning code base for the project, so they are familiar with many of the necessary tools. Additionally Savannah has some experience web scraping by working with Selenium, which is the primary tool we will be using to collect data from the web. Nick does not have experience with the specific Python libraries we will be leveraging, but does have some experience with the Python language.

The roles within the group will be highly flexible, with Isaac acting as the team leader. Since we're a 4 person team, each member will likely be required a number of different types of tasks throughout the project. Roles will be assigned during each sprint so that each member has a specific piece of the project to work on at all times.

Project Management

Completion of a sentiment analysis interface as well as the systems for acquiring and creating the data should be doable within the few months we will be in this class, especially since some of the desired functionality has already been implemented. We plan to meet weekly face to face to discuss individual progress and help each other if someone is stuck. Our main goals are monthly, as we have three main goals and this would put us at having our project completed in late April, however since our goals are not completely linear, we plan to be working on multiple goals at a time based on our skills. Our tentative schedule is to have the web scraping script done by mid-February, the classification algorithm done by mid-March, and UI finished by mid-April, which will leave us enough time to put together a presentation for the end of this course. This time table will become more specific when we've had another chance to meet and see how long individuals think more specific tasks will take. However as a very tentative week-by-week schedule we have, with an estimated 12 weeks :

Week 1 : Meet and determine individual strengths and determine where team members will begin working on the project

Week 2 : Have beginnings of web scraping script - begin work on classification algorithm

Week 3: Have websites / articles in mind to be used for web scraping - have metrics in mind for negative / positive opinions of cryptocurrencies

Week 4: Have scraping script mostly finished - continue to look at how to analyze positive v negative articles

Week 5: Have all data from web scraping - begin fully working on classification

Week 6: Work on classification

Week 7: Work on classification - begin brainstorming ideas for UI

Week 8: Work on classification - brainstorm UI

Week 9: Be wrapping up classification - beginnings of UI

Week 10: Finished classification algorithm - fully switch to UI

Week 11: Finish up UI

Week 12: Wrap up / create presentation

As far as we are aware there are no constraints to this project. As our resources are online articles that the public views and cryptocurrency prices, we should have no trouble accessing these resources. If the full scope of this project cannot be completed, we will still have useful metrics on how people are currently feeling about cryptocurrencies. Even if we end up unable to create a UI that shows prices correlated to historical sentiment of the currency, we will still be able to show trends in sentiment over time, as well as current sentiments.

Reflection

With our team members strengths regarding the software we plan to use, as well as already having the beginnings of code, we think that this project is completely doable within the time frame of this course. We also believe that with the rise in both popularity and number of cryptocurrencies, this tool would be very interesting to create and to analyze. While we believe that there will be a correlation between sentiment and price, that's something we would like to be able to show concretely, and also believe it would also be interesting to find out that there is no correlation. This is something that may change the scope of our project. If we finish classifying articles and looking at prices relative to sentiment and find that there is no correlation, than creating a user interface to show the current sentiment along with prices becomes essentially worthless. If this happens, we hope to have time to begin looking in to other factors that may contribute to cryptocurrency prices. While we would probably not have time to create a UI for this case, since it will be time consuming to find what does have a correlation, this would also be valuable and interesting information.